Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.7 Release Notes

Rev 4.17.15.16

NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT (“PRODUCT(S)”) AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES “AS-IS” WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCT(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Mellanox Technologies
350 Oakmead Parkway Suite 100
Sunnyvale, CA 94085
U.S.A.
www.mellanox.com
Tel: (408) 970-3400
Fax: (408) 970-3403

Table of Contents

List Of Tables

Release Update History

Table 1 - Release Update History

Release	Date	Description
Rev 4.17.15.16	March 25, 2019	Initial release of this MLNX-NATIVE-ESXi-ConnectX-4/ ConnectX-4 Lx/ConnectX-5 driver version

1 Overview

These are the release notes of Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.7. Mellanox ConnectX-4/ConnectX-5 NATIVE ESXi Driver for VMware vSphere 6.7 supports the following uplinks to servers

Table 2 - Supported Uplinks to Servers

Version	OS	Uplink Speed
4.17.15.16	ESXi 6.7, ESXi 6.7 U1	10/25/40/50/100GbE

1.1 Content of MLNX-NATIVE-ESX Driver Package

The MLNX-NATIVE-ESX driver package is distributed as an offline bundle (.zip file) and contains:

- **ESXi 6.7:**

  MLNX-NATIVE-ESX-ConnectX-4-5_4.17.15.16-10EM-670.0.0.8169922.zip - Hypervisor bundle for ESXi 6.7 contains the following kernel modules:

  - nmlx5_core
  - nmlx5_rdma

1.2 Supported HCAs Firmware Versions

MLNX-NATIVE-ESX Rev 4.17.15.16 supports the following Mellanox Ethernet HCA:

Table 3 - Supported HCAs Firmware Versions

HCAs	Recommended Firmware Rev.	Additional Firmware Rev.
ConnectX-4	12.24.1000	12.23.1020
ConnectX-4 Lx	14.24.1000	14.23.1020
ConnectX-5	16.24.1000	16.23.1020
ConnectX-5 Ex	16.24.1000	16.23.1020

For the latest firmware versions, visit:

- http://www.mellanox.com/page/products_dyn?product_family=29

1.3 Tested Hypervisors in Paravirtualized and SR-IOV Environments

Table 4 - Tested Hypervisors in Paravirtualized and SR-IOV Environments

Tested Hypervisors	HCAs	Guest Operating System
SR-IOV	ConnectX-4/ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex	Windows Server 2012 R2
		RedHat 6.5
		RedHat 6.6
		RedHat 7.1
		RedHat 7.2
		RedHat 7.3
		RedHat 7.5
Paravirtualized[a] (Ethernet Only)	ConnectX-4/ConnectX-4 Lx ConnectX-5/ConnectX-5 Ex	RedHat 7.5

a. Paravirtualized RDMA is only tested Linux OSes.

2 Changes and New Features in Rev 4.17.15.16

Table 5 - Changes and New Features

Feature/Change	Description
Enhanced Network Stack (ENS)	Enhanced data path is a networking stack mode, which when configured provides superior network performance. It is primarily targeted for NFV workloads, which requires the performance benefits provided by this mode. ENS utilizes DPDK Poll Mode driver model and significantly improves packet rate and latency for small message sizes. This feature is compliant with NSX-T version 2.3.1.0.0.11294271. **Note:** This driver version maintains support of all previous Ethernet functionalities, and can operate in both ENS and slow-path network stack mode, based on the DVS configuration.

3 Known Issues

The following is a list of general limitations and known issues of the various components of this MLNX-NATIVE-ESX release.

Table 6 - Known Issues (Sheet 1 of 3)

Internal Ref.	Description
1671303	**Description:** A PSOD may occur during vMotion over ENS VMK. This issue is pending VMWare investigation.
	Workaround: N/A
	Keywords: ENS, vMotion
	Discovered in Version: 4.17.15.16
1668029	**Description:** There is no Geneve traffic on NSX-T 2.3 when DVS is in standard (non-ENS) mode.
	Workaround: Use NSX-T 2.2 for this scenario or use DVS in ENS mode with NSX-T 2.3
	Keywords: ENS, Geneve
	Discovered in Version: 4.17.15.16
1677627	**Description:** IPv6 as inner packet is not supported by VMWare yet. This capability will be added in next NSX-T release.
	Workaround: N/A
	Keywords: ENS, IPv6, Geneve
	Discovered in Version: 4.17.15.16
1682956	**Description:** During ENS uplink detachment from the ENS DVS, the below error message regarding the queue still being allocated or that the requested queue is not in use may appear. *"Driver covers for OS issue and the messages are for information only."*
	Workaround: N/A
	Keywords: ENS, queue
	Discovered in Version: 4.17.15.16
1712298	**Description:** Live unload of the driver is not supported. Doing so may cause a PSOD if the max_vfs parameter is set.
	Workaround: N/A
	Keywords: Driver load
	Discovered in Version: 4.17.15.16

Table 6 - Known Issues (Sheet 2 of 3)

Internal Ref.	Description
1446060	**Description:** Although the max_vfs module parameter range is "0-128", due to firmware limitations, the following are the supported VFs per single port devices: • ConnectX-4: up to 127 • ConnectX-5: up to 63
	Workaround: N/A
	Keywords: SR-IOV, VFs per port
	Discovered in Version: 4.17.14.2
1340275	**Description:** ECN tunable parameter `initialAlphaValue` for the Reaction Point protocol cannot be modified.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli
	Discovered in Version: 4.17.13.8
1340255	**Description:** ECN statistic counters `accumulatorsPeriod` and `ecnMarkedRocePackets` display wrong values and cannot be cleared.
	Workaround: N/A
	Keywords: nmlx5 ecn nmlxcli
	Discovered in Version: 4.17.13.8
-	**Description:** The hardware can offload only up to 256B of headers.
	Workaround: N/A
	Keywords: Hardware offload
	Discovered in Version: 4.17.13.8
781277	**Description:** The "`esxcli network sriovnic vf stats`" command is not supported. When running this command on a vmknic, a failure message is displayed.
	Workaround: N/A
	Keywords: esxcli SR-IOV
	Discovered in Version: 4.17.13.8
858972	**Description:** Setting the "Allow Guest MTU Change" option in vSphere Client is currently not functional. Although guest MTU changes in SR-IOV are allowed, they do not affect the port's MTU and the guest's MTU remains the same as the PF MTU.
	Workaround: N/A
	Keywords: MTU, SR-IOV
746100	Description: The '`esxcli mellanox uplink link info -u <vmnic_name>`' command reports the 'Auto negotiation' capability always as 'true'.
	Workaround: N/A
	Keywords: 'Auto negotiation' capability

Table 6 - Known Issues (Sheet 3 of 3)

Internal Ref.	Description
1068621	Description: SMP MADs (ibnetdiscover, sminfo, iblinkinfo, smpdump, ibqueryerr, ibdiagnet and smpquery) are not supported on the VFs.
	Workaround: N/A
	Keywords: SMP MADs
778371	**Description:** Wake-on-LAN does not notify when invalid parameters are provided.
	Workaround: N/A
	Keywords: WoL
778572	**Description:** Nested ESXi might not function properly.
	Workaround: N/A
	Keywords: Nested ESXi
765008	**Description:** Device RSS fails to hash traffic to sufficient RX rings with Broadcast traffic.
	Workaround: N/A
	Keywords: RSS, RX rings
852883	**Description:** In stress condition ‘Watchdog’ may appear, leading to uplink going up and down.
	Workaround: N/A
	Keywords: uplink, watchdog

4 Bug Fixes History

Table 7 lists the bugs fixed in this release.

Table 7 - Fixed Bugs List

Internal Ref.	Description
1358381	**Description:** Fixed an issue that prevented ESXi from being discovered via the CDP protocol on ConnectX-4 Lx adapter cards.
	Keywords: CDP protocol, ConnectX-4 Lx
	Discovered in Release: 4.17.13.8
	Fixed in Release: 4.17.14.2

5 Change Log History

Table 8 - Change Log History (Sheet 1 of 3)

Feature/Change	Description
Rev. 4.17.14.2	
Virtualization	Removed a VF driver limitation. Now the driver can support the maximum number of VFs supported by the firmware.
Bug Fixes	See Section 4, “Bug Fixes History”, on page 12
Rev. 4.17.13.8	
Explicit Congestion Notification (ECN)	Explicit Congestion Notification (ECN) is an extension to the Internet Protocol and to the Transmission Control Protocol. ECN allows end-to-end notification of network congestion without dropping packets. To configure ECN behavior, download the nmlxcli tool from the Mellanox site. For further information, refer to the User Manual section *Explicit Congestion Notification (ECN)*.
Dynamic RSS	Improves network performance by allowing OS Load Balancer better RSS RX queue utilization during heavy traffic of the same type. For further information, refer to the User Manual section *Dynamic RSS*.
Multiple RSS Engines	Improves network performance by exposing multiple RSS RX queues to the hypervisor network stack. For further information, refer to the User Manual section *Multiple RSS Engines*.
Packet Capture Utility	Packet Capture utility duplicates all traffic, including RDMA, in its raw Ethernet form (before stripping) to a dedicated "sniffing" QP, and then passes it to an ESX drop capture point. It allows gathering of Ethernet and RoCE bidirectional traffic via pktcap-uw and viewing it using regular Ethernet tools, e.g. Wireshark To enable/disable packet capture, download the nmlxcli tool from the Mellanox site. For further information, refer to the User Manual section Packet Capture Utility.
SR-IOV max_vfs module parameter Type Modification	Changed the type of the SR-IOV max_vfs module parameter from a single integer value to an array of unsigned integers. For further information, refer to the User Manual.
InfiniBand SR-IOV	Enables the creation of InfiniBand virtual functions, allowing the guests to operate over an InfiniBand fabric.
DCBX Negotiation Support for PFC	PFC port configuration can now be auto-negotiated with switches that support the DCBX protocol.
ESXi CLI	Added ESXi CLI support for ESXi 6.7
Adapter Cards	Added support for ConnectX-5/ConnectX-5 Ex adapter cards. **Note:** ConnectX-5/ConnectX-5 Ex cards are currently at beta level.

Table 8 - Change Log History (Sheet 2 of 3)

Feature/Change	Description
Geneve Stateless Offload	Geneve network protocol is encapsulated into IP frame (L2 tunneling). Encapsulation is suggested as a means to alter the normal IP routing for datagrams, by delivering them to an intermediate destination that would otherwise not be selected based on the (network part of the) IP Destination Address field in the original IP header.
Remote Direct Memory Access (RDMA)	Remote Direct Memory Access (RDMA) is the remote memory management capability that allows server-to-server data movement directly between application memory without any CPU involvement. **Note:** It is recommended to use RoCE with PFC enabled in driver and network switches. For how to enable PFC in the driver see section *Priority Flow Control (PFC)* in the User Manual.
Set Link Speed	Enables you to set the link speed to a specific link speed supported by ESXi. For further information, see the User Manual section *"Set Link Speed"*.
Priority Flow Control (PFC)	Applies pause functionality to specific classes of traffic on the Ethernet link. For further information, see the User Manual section *"Priority Flow Control (PFC)"*.
NetQ RSS	Allows the user to configure multiple hardware queues backing up the single RX queue. NetQ RSS improves vMotion performance and multiple streams of IPv4/IPv6 TCP/UDP/IPSEC bandwidth over single interface between the Virtual Machines. For further information, see the User Manual section *"NetQ RSS"*.
Default Queue RSS (DRSS)	Allows the user to configure multiple hardware queues backing up the default RX queue. DRSS improves performance for large scale multicast traffic between hypervisors and Virtual Machines interfaces. For further information, see the User Manual section *"Default Queue Receive Side Scaling (DRSS)"*.
SR-IOV	Single Root IO Virtualization (SR-IOV) is a technology that allows a physical PCIe device to present itself multiple times through the PCIe bus.
	Support for up to 8 ConnectX-4 ports and up to 16 VFs. For further information, refer to the User Manual
RX/TX Ring Resize	Allows the network administrator to set new RX\TX ring buffer size.
VXLAN Hardware Stateless Offloads for ConnectX®-4	VXLAN hardware offload enables the traditional offloads to be performed on the encapsulated traffic.
NetDump	Enables a host to transmit diagnostic information via the network to a remote netdump service, which stores it on disk. Network-based coredump collection can be configured in addition to or instead of disk-based coredump collection.
NetQueue	NetQueue is a performance technology in VMware ESXi that significantly improves performance in Ethernet virtualized environments.
Wake-on-LAN	Allows a network administrator to remotely power on a system or to wake it up from sleep mode

Table 8 - Change Log History (Sheet 3 of 3)

Feature/Change	Description
Hardware Offload	• Large Send Offload (TCP Segmentation Offload) • RSS (Device RSS)
Hardware Capabilities	• Multiple Tx/Rx rings • Fixed Pass-Through • Single/Dual port • MSI-X
Ethernet Network	• TX/RX checksum • Auto moderation and Coalescing • VLAN stripping offload